DETECTION OF MEDICAL IDENTITY THEFT IN MEDICAL IMAGES THROUGH DIGITAL WATERMARKING

BY

MUKTAR YAHUZA

A dissertation submitted in fulfilment of the requirement for the degree of Master in Computer and Information Engineering

Kulliyah of Engineering
International Islamic University Malaysia

JANUARY 2015

ABSTRACT

The risk of medical identity theft is increased rapidly due to the advancement in technology. Accordingly, the healthcare system security and privacy becomes an important issue. This is because any alteration to medical information may cause a devastating effect to the patient. In this work, a digital watermarking that detects any attempt of altering patient data is proposed. The discrete wavelet transform of the 8 x 8 non-overlapping blocks of the medical test image is generated prior to the embedding process. A 64-bits binary equivalent of digit numbers representing patient's entry date and file ID used as the watermark is embedded inside the corresponding patient's medical image by quantizing the coefficient of the highest frequency components of each block. After the medical image is transferred to its destination, the watermark is extracted and compared to the original watermark for authentication. The average value of the peak signal to noise ratio performance metric shows that the level of imperceptibility of the technique, was found to be 82.88 *d*B, and also the average values of the mean square error, and that of structural similarity showed that the level of distortion of the watermarked image was found to be $7.9e^{-4}$ and 0.9112 respectively. A number of attacks which include JPEG compression, Filtration, Gaussian noise, Salt and pepper noise, and contrast enhancement were applied to the watermarked image. The proposed algorithm enables quick and excellent detection capability of any modification done to the medical image.

خلاصة البحث

إن خطورة التعدي على بيانات أي مريض قد ارتفع بشكل سريع نظراً لتقدم التقنيات الحديثة. و بهذا الخصوص أصبح نظام متابعة المريض من الناحية الأمنية و الخصوصية مهم جداً لأن أي تغيير في بيانات المريض الطبية قد يؤدي إلى نتيجة مدمرة على المريض. في هذا العمل تم اقتراح علامة مائية رقمية لاكتشاف أي محاولة لتغيير بيانات المريض. تحويل الإشارة الصغيرة الى 8×8 كتل غير متداخلة تم توليدها للصورة الطبية المستخدمة في الاختبار قبل عملية الإدخال. التمثيل الثنائي من 64 بت لبيانات المريض المدخلة وملف التعريف الخاص به استخدمت كعلامة مائية وتم إدخالها داخل صورة المريض الطبية بواسطة تكميم معاملات أعلى مركبات ترددية لكل كتلة. بعد إرسال صورة المريض الطبية إلى الوجهة المقصودة، تم استخلاص العلامة المائية ومقارنتها بالعلامة المائية الأصلية لغرض التحقق. معيار قيمة متوسط قمة الإشارة إلى الى الضوضاء، الذي يظهر مستوي عدم الإدراكية لهذه التقنية، كان 82.88 dB. بالإضافة إلى أن متوسط الخطأ المربع، الذي يظهر مستوى التشويه في العلامة المائية، كان 0.9112 ,7.9e-4. العديد من الهجمات التي تتضمن ضغط JPEG، عمليات الفلترة، الضوضاء الجاوسية، ضوضاء الملح والورق بالإضافة إلى تحسين التغاير تم تطبيقها على صورة العلامة المائية، النموذج المقترح كان قادرا على اكتشاف كل واحد من هذه الهجمات لذا فالخوارزمية المقترحة تقدم قدرات استكشافية سريعة وسهلة لأي عملية تعديل تمت على الصورة الطبية.

APPROVAL PAGE

I certify that I have supervised and read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Master of Science in Computer and Information Engineering

Rashidah Funke Olanrewaju
Supervisor

Othman O. Khalifa
Co-Supervisor

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Master of Science in Computer and Information Engineering

Khairul Azami Sidek
Examiner

Teddy Surya Gunawan
Examiner

This dissertation was submitted to the Department of Electrical and Computer Engineering and is accepted as a fulfilment of the requirement for the degree of Master of Science in Computer and Information Engineering

Othman O. Khalifa
Head, Department of Electrical
and Computer Engineering

This dissertation was submitted to the Kulliyyah of Engineering and is accepted as a fulfilment of the requirement for the degree of Master of Science in Computer and Information Engineering

Md. Noor B. Salleh
Dean, Kulliyyah of Engineering

DECLARATION

I hereby declare that this thesis is the result of my own investigations, except where otherwise stated. I also declare that it has not been previously or concurrently submitted as a whole for any other degree at IIUM or other institutions.

Muktar Yahuza

Signature _____ Date_____

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

DECLARATION OF COPYRIGHT AND AFFIRMATION OF FAIR USE OF UNPUBLISHED RESEARCH

DETECTION OF MEDICAL IDENTITY THEFT IN MEDICAL IMAGES THROUGH DIGITAL WATERMARKING

Affirmed by Muktar Yahuza

Signature _____ Date_____

ACKNOWLEDGEMENT

Assalamu alaikum warahmatullahi wabarakatuhu,

Alhamdulillah. All praises is due to Almighty Allah (S.W.T), the most Gracious and the most Merciful, Lord of the universe. Peace and blessing be upon our noble prophet Muhammad (S.A.W). Alhamdulillah, I am very grateful that Allah S.W.T. has given me the strength, patience, courage and determination in compiling this thesis.

First and foremost, I would like to express my appreciation to my supervisor, Dr. Rashidah Funke Olanrewaju for her never ending support, guidance, cooperation and patience in guiding me to complete my dissertation. This dissertation would not be complete without her motherly and professional guidance and supports. I am indebted to her for her patience in correcting and improving my draft before I can come with the final version of this dissertation. I will also like to thank my co-supervisor, Prof. Othman O.khalifa for his endless support and guidance.

Secondly, special thanks to all lecturers of Department of Electrical and Computer Engineering, Kulliyyah of Engineering of International Islamic University Malaysia (IIUM) as these are the people who are responsible of giving me knowledge and who taught me so well throughout the journey. My gratitude also goes to the administration staffs that always assist me at the Kulliyah of Engineering Post Graduate Research office especially regarding the dissertation procedure. May Allah reward them for their kindness and assistance throughout my study period in IIUM.

Thirdly, I would like to thank my beloved brother, Alhaji Abdulhadi Yahuza, parents Alhaji Yahuza and Hajiya Umma, my fiancée Saeeda, my family members and friends for their prayers supports that were given to me in the course of my study. Obstacles encountered would not be easily overcome without their supports and prayers.

Above all, I would like to thank our beloved Government of Kano State of Nigeria, His Excellency Engr. Dr. Rabiu Musa Kwankwaso for his forward looking concern and support to further my post graduate studies. Equally, my sincere gratitude goes to Kano State Scholarship staff may Almighty Allah continue to guide them.

Finally, a big thank to everyone who have helped me directly or indirectly in the completion of this dissertation, I sincerely appreciate everything that was given to me, and the blessings that were bestowed to me to have good people like all of you in my life. May Allah SWT reward you and your family for the aforementioned endeavors. Honestly, I can never thank you enough for your kindness and assistance. Last but not the least, I presented this dissertation as a symbol of gratitude for everyone, and as the celebration of knowledge that we have gained alone the ride, in hopes that it will benefit the Ummah.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

MIDT	Medical Identity Theft ITRC Identity theft resource Centre
MRI	Magnetic Resonance Imaging
CT	Computed Tomography images
HIPAA	Health Insurance Portability and Accountability Act
EMR	Electronic Medical Records
DTS	Department of Technology Services
HIS	Health Information Systems
BCBST	Blue Cross Blue Shield of Tennessee
HHS	Health and Human Services
AHRQ	Agency for Healthcare Research and quality
EPR	Electronic Patient Records
DWT	Discrete Wavelet Transform
MATLAB	Matrix Laboratory
PSNR	Peak Signal to Noise Ratio
MSE	Mean Square Error
SSIM	Structural Similarity
LSB	Least Significant Bit
RSA	Rivest Shamir Adleman
MD5	Message Digest 5
AES	Advanced Encryption Standard
NSCT	Non-Subsampled Contourlet Transform
ROI	Region of Interest
RONI	Region of Non-Interest
KLT	Karhumen Loeve transform
ECE	Error Correcting Codes
DCT	Discrete Wavelet Transform
IWT	Integer Wavelet Transform

MSB	Most Significant Bits
CVNN	Complex Valued Neural Network
FFT	Fast Fourier Transform Function
DE	Expansion Technique
EXIF	Exchangeable Image File Format
JPEG	Joint Photographic Expert Group
CWT	Continuous wavelet transform
DWT	Discrete wavelet transform
FWT	Fast wavelet transform
WPD	Wavelet packet decomposition
SWT	Stationary wavelet transform
FRWT	Fractional wavelet transform
SHVS	Superior Human Visual system
IDWT	Inverse discrete wavelet Transform
Bpp	Bit per Pixel
CCITT	Comite Consultatif International Telegraphique et Telephonique
ISO	International Organization for Standardization

LIST OF SYMBOLS

Log_{10}	Log to base 10
MAX_f	Peak value of the image signal
$f(i, j)$	Pixel of the medical image
$W(i, j)$	Binary pixel of the watermark
Q_e	Quantization to the nearest even number
Q_o	Quantization to the nearest odd number
Δ	The scaling quantity which is the same as the quantization step used to quantize either to an even or odd number
H_k	The representation of the predefine coefficient in each 8*8 sub blocks

CHAPTER ONE

INTRODUCTION

1.1 PREAMBLE

In the modern world of today, the change from the oldest and manually based health record system to the electronic healthcare record is being given much interest. With the development of modern healthcare services, the long-established way of achieving health care service where the patient must meet with the health personnel is reduced. The use of electronic information and communication technologies to provide healthcare to patients who are separated by distance is referred to as Telemedicine (Liqiong & Marshal, 2002). Telemedicine plays a significant role in providing services by improving value, justice and contact through connecting healthcare facilities and medical personnel with the patient, as well as eliminating geographical barriers associated with the existing medical practice. Although the approach provides many benefits for healthcare data delivery, however, there are a number of security and privacy implications that must be explored in order to promote and maintain the ultimate medical ethics. The main threat is from the emergent security issues which include medical identity theft (MIDT) and medical information attacks.

MIDT is seen to be more evolving among the available medical threats (Ponemon, 2013). Most of the time people wonder why medical information is of interest to those who are not the eligible owners of the information. Medical records contain two categories of information; personal information of the patient which includes name, date of birth, social security number and the other category is the financial record which involves all the banking information of the patient such as credit card number and insurance card number and etc. Hence, due to the monetary

information involved in telemedicine, medical identity theft is seen to be increasing rampantly.

Medical identity theft can be referred as an act in which an individual uses another person's medical identity without his/her knowledge and consent to obtain medical services or benefits. The victims of medical identity theft are the patients, and other health personnel such as the physicians and nurses. Hackers may steal medical records of a patient and sell them to others in order to; obtain healthcare services of the victims, obtain pharmaceuticals or other medical equipment, obtain governmental benefits; bill the healthcare plan, insurance company or government program, or they may alter the record in order to be claiming the patient's future medical benefits. Sometimes, MIDT causes the victims not only to lose their financial assistances, but also to suffer from wrong diagnosis which may lead to their death (Figg & Kam, 2011).

Medical identity theft is seen to be increasing every year. The most recent survey conducted by Ponemon Institute in September 2013 revealed that 1.52 million Americans were affected by MIDT in the year 2012, which rises to 1.84 million in the year 2013. This is almost a 32% increase in just one year (Ponemon, 2013). Furthermore, Ponemon Institute also gave out a report figuring that the growing of medical identity theft is estimated to have affected as many as 1,800,000 people in the year 2013. These victims had their information illegally used to receive medical care, benefits or insurance. Wrong information in health files can lead to negative consequences to affected victims. Fabricated entries on medical files can hinder an individual's medical coverage and, in some instances, make them uninsurable (e.g. having a disease that is not yours). Sometimes, it can lead victims to be unemployable, e.g. if it contain psychiatric history. This may not be discovered until incorrect

medical treatment or outstanding bills appeared in the victim's file. Unlike credit card report, patients do not have the same rights to correct errors in their medical histories, and also they do not have the right to receive a free copy of their medical file as compared to a credit card report. In an event reported, a teenager was denied the opportunity to donate blood because Red Cross marked her social security number as belonging to a person who was tested positive for HIV (Rick & Christine, 2012).

Eva Velasquez, President and CEO of the Identity Theft Resource Centre (ITRC) has reported to have said; "MIDT has dramatically increased due to the encouragement given by the HIPAA/HITECH policies to use electronic health records that were valued to different data thieves, and also due to the potential increase of hackers who steal medical information and sell them in the black market" (Identity Theft Resource Centre [ITRC], 2013). She also added that; "Because MIDT is already a sophisticated growing problem, the ITRC believes that it is important to educate the public about the consequence of it, as well as the risks associated with it" (ITRC, 2013).

A lot of instances of medical identity theft have occurred. Although among such instances, few occurred accidentally, for example, loss or theft of computers belonging to health organizations. However, most of the instances happened intentionally. Intentional leak of healthcare data occur when dishonest employee disclose the medical data to external parties for the purpose of monetary gain. A typical example is the case of a Cleveland Clinic of U.S (Health Insurance Portability and Accountability Act [HIPAA], 2013). An employee in the clinic sold a patient's healthcare record to her cousin, who ran medical claims of a company. He then went ahead and filed many fake medical claims for monetary purposes. The consequences of his action led to financial loss to the victim which at the end tarnished the

confidence level of patient to medical professional in e-healthcare system. Furthermore, the effectiveness of electronic health records are questionable to the sharing of Electronic Medical Records (EMR) among business partners and other entities, makes the crime to continually increase. Recent incident of MIDT is when a hacker from Eastern Europe illegally accessed Utah Department of Technology Services (DTS) server containing patient's social security numbers and data on children's health plans due to a weak password protection scheme. The breach involved 780,000 individuals both medical patients as well as recipients of children's health insurance plan stolen from the server (Olanrewaju, Nor'ashikin, Othman, & Azizah, 2013). Cyber-attacks on patient's EMR and health information systems (HIS) can lead to severe consequences like patient identity disclosure, embarrassment, privacy violation and in the worst case, integrity violation resulting in patient's death (Das & Mukhopadhyay, 2011). In March 2012, Blue Cross Blue Shield of Tennessee (BCBST) agreed to pay U.S. Department of Health and Human Services (HHS) $1.5 million for failure in providing the adequate security measures which as a result allowed over 57 unencrypted hard drives which contains secretive health information to be stolen from its facility (Health and Human services [HHS], 2012).

Numerous protective, limiting, and corrective measures have been employed to minimize the damage caused to telemedicine information due to medical identity theft. Among the measures are; forming and enactment of the Health Insurance Portability and Accountability Act HIPAA of 1996. It was intended at developing criteria and requirements for the transmission and maintenance of medical information electronically (HIPAA, 2013). A similar regulatory body is the creation of Agency for Healthcare Research and quality (AHRQ), which is a U.S. federal agency, under

health and human services working to improve the quality, effectiveness and safety of healthcare (Lockwood W., 2013).

Likewise, various technological techniques apart from the law enactment have been developed to reduce the effect of MIDT. Among the technological measures are cryptography, steganography, and digital watermarking (Olanrewaju et al., 2013).

Although cryptography is used in protecting medical information from theft and attack, however, due to the fact that the technique loses its protection once it is decrypted, medical data cannot be fully protected well with this method (Olanrewaju et al., 2013). Also, steganography essentially exploit human perception, and due to the fact that human senses are not trained to look for files that have information hidden inside them, the technique prevents any human intervention to the information (Chandramouli, Rajarathnam & Nasir, 2003). However, hacker (third party) employed computer system to detect and attack steganography information. Hence the technique failed to serve as suitable means of curving security and privacy attacks involved in telemedicine system.

Digital watermarking in contrast to encryption and steganography has been used more frequently by various researchers as a mean of addressing the modern health management data issues, including enhanced security of sensitive data, and its source authentication (Olanrewaju et al., 2013). This is because of its ability to fight against digital piracy, authenticates and verifies the integrity of digital information. In addition, the digital watermark's vital features such as, imperceptibility, inseparability of the content from the watermark, and its essential ability to undergo same transformation experienced by the host signal make it preferable over the methods given above of protecting medical data integrity (Olanrewaju et al., 2013).

1.2 OVERVIEW OF DIGITAL WATERMARKING

Digital watermarking which is a term used back from paper watermarking, has the additional concept of resilience against any attempts to remove the hidden data when compared to the other technological measures. It is defined as the practice of invisibly altering a given data (such as multimedia) to embed a message about that data (Ingemar, Miller, Bloom, Fridrich, & Kalker, 2008). It consists of embedding and extracting blocks as illustrated in Figure 1.1. The embedding block takes two inputs. One is the information to be embedded, i.e. watermark or the secret message, and the other is the host data in which the watermark is to be embedded to it (Ingemar et.al, 2008).

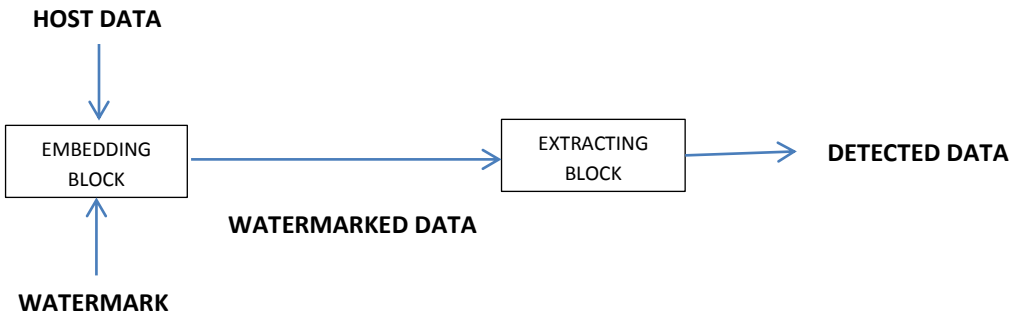

Figure 1.1: A typical watermarking system

Digital watermarking is agreed by various researchers to be sufficient in providing better medical image security (Olanrewaju et al., 2013). However, the originality of the information in the image must be preserved to avoid any performance loss for the medical professional viewers. The loss of information is considered to be the main factor that can prevent proper diagnoses or treatment, which can lead to a lot of damages. Very few watermarking techniques for detecting medical

identity theft are available, and most of them introduce distortion, visual artifact or noise, and sometimes even loss of information during the watermarking implementation. In medical issues, loss of information, distortion, and addition of artifact to the patient image record are not acceptable. Thus, for any successful telemedicine watermarking technique, the above limitations have to be mitigated.

Among the previous techniques, some embed the watermark directly into the medical image which resulted in distortion of the image. Applying the other one, i.e. frequency domain watermarking result into having real and imaginary parts of the host image, and the effect may cause the original image to be distorted. This is because only the real or imaginary part of the original image is considered during the embedding process. Any distortion or addition of artifacts during the watermarking process will cause difficulty and sometime even hinder the detection of forged watermarks introduced by hackers during the attempt of medical identity theft.

Therefore, a strong recommendation in telemedicine field for having an optimal distortion free watermarking technique is required. In this research, an optimal digital watermarking technique will be used to detect any attempt to alter, steal, or delete medical record by a third party. This work will help in detecting medical identity theft earlier before it start affecting the victim, thus overcoming the late notification of the threat. With this proposed technique, it is hoped that our study will contribute towards the development in the telemedicine field.

1.3 OVERVIEW OF TELEMEDICINE

Telemedicine can be defined as the practice of medicine when the doctor and the patient are widely separated using mutual speech and graphic communication (e.g.

satellite, computer, and television) (Sood, Mbarika, Jugoo, Dookhy, Doarn, Prakash, & Merrell, 2007). It can also be defined as the exchange of medical information from one location to another with the use of modern telecommunication and information technology (Perednia, Douglas & Ace, 1995).

Telemedicine is the use of modern technology to provide healthcare to participants who are separated by distance (Liqiong & Marshal, 2002). "Tele" comes from a Greek word meaning "at a distance", therefore telemedicine is a field where all the consultation is performed at a distance (Giakoumaki, Perakis, Tagaris, & Koutsouris, 2009). With the development of modern healthcare services, the traditional face-to-face method between the patient and health personnel is not necessary. Thus, with the modern development, telemedicine applications are becoming popular. Telemedicine plays a significant role in improving the healthcare sector by connecting healthcare facilities and medical personnel with the patient, as well as eliminating geographical barriers between them.

Figure 1.2 shows a schematic diagram of a typical telemedicine project. Modern telecommunication technologies and information systems are being combined and implemented in telemedicine field to store patient's data such as blood test result from medical lab, and x-rays, radiographic and urology system information in digital format called the electronic patient record (EPR) that allows adequate medical support delivered to the patient who is located in an entirely different area, which will be accessible by the doctors through the internet (Furuie, Rebelo, Moreno, Santos, Bertozzo & Motta, 2007).

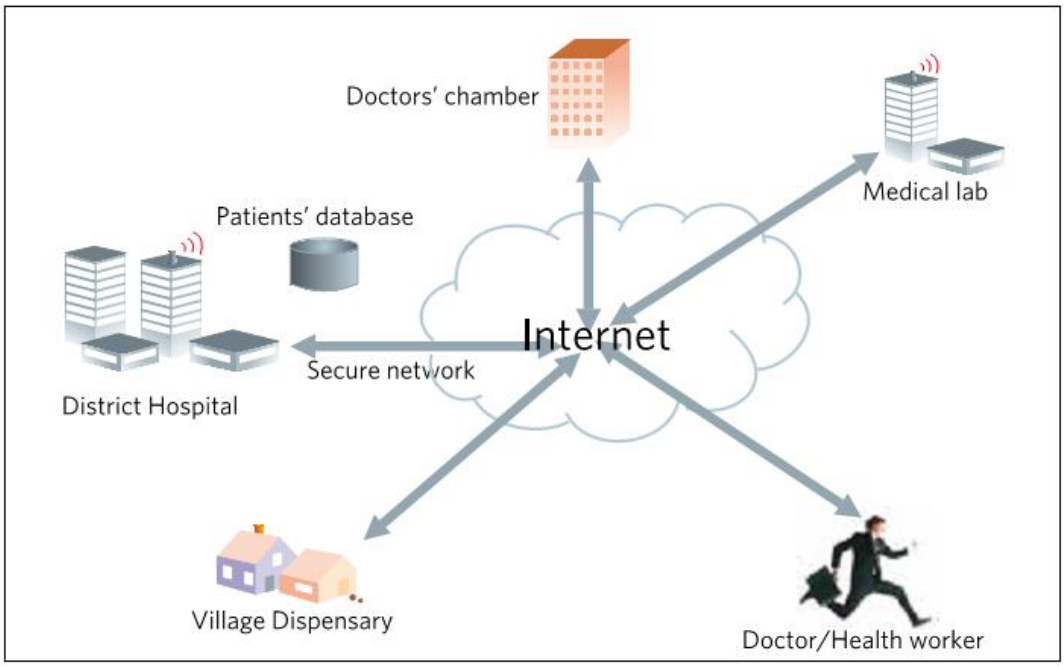

Figure 1.2: Schematic diagram of a telemedicine project (Source: Das & Mukhopadhyay, 2011)

With the use of EPR, health care records are now retrieved in real-time irrespective of the location of the patient. As such, medical personnel, as well as insurance companies, and patients can be able to access it over the Internet (Meingast, Roosta, & Sastry, 2006). The potential loss of data associated with the medical data is now solved with the use of EPRs, as medical data can now be backed up. A local database which is connected to the internet is used to collect information of patient at a given location, as illustrated in Figure 1.2. This allows the medical personnel at a given hospital to access a patient's information from entirely different hospital, possibly located at different continent (Simon, 2000).

In a telemedicine system, sensor network is used in combination with EPR for distant patient monitoring, diagnosis, and consultation (Brown, 2005). With the improvement of sensor networks, patients can be monitored right away from their home synchronously. Figure 1.3 illustrates a remote patient monitoring system used in